Cognitive ability and the effect of strategic uncertainty

Nobuyuki Hanaki, Nicolas Jacquemet, Stéphane Luchini, Adam Zylbersztejn

Cognitive ability and the effect of strategic uncertainty*

Nobuyuki Hanaki† Nicolas Jacquemet‡ Stéphane Luchini§ Adam Zylbersztejn¶

October 2015

Abstract

How is one's cognitive ability related to the way one responds to strategic uncertainty? We address this question by conducting a set of experiments in simple 2×2 dominance solvable coordination games. Our experiments involve two main treatments: one in which two human subjects interact, and another in which one human subject interacts with a computer program whose behavior is known. By making the behavior of the computer perfectly predictable, the latter treatment eliminates strategic uncertainty. We find that subjects with higher cognitive ability are more sensitive to strategic uncertainty than those with lower cognitive ability.

Keywords: Strategic uncertainty, Bounded rationality, Robot, Experiment

JEL Classification: C92, D83.

*This project is partly financed by JSPS-ANR bilateral research grant "BECOA" (ANR-11-FRJA-0002). Part of this research was performed within the framework of the LABEX CORTEX (ANR-11-LABX-0042) of Université de Lyon, within the program "Investissements d'Avenir" (ANR-11-IDEX-007) operated by the French National Research Agency (ANR). Ivan Ouss provided efficient research assistance. We thank Juergen Bracht, Colin Camerer, Guillaume Fréchette, Haoran He, Asen Ivanov, Frédéric Koessler, Rosemarie Nagel, Ariel Rubinstein, Jason F. Shogren, Jean-Marc Tallon, Antoine Terracol and Marie-Claire Villeval for their comments. Hanaki and Jacquemet gratefully acknowledge the *Institut Universitaire de France*. Luchini thanks the School of Business at the University of Western Australia for hospitality and support. A major part of this work was conducted while Hanaki was affiliated with Aix-Marseille University (Aix-Marseille School of Economics, AMSE) and Jacquemet was affiliated with Université de Lorraine (BETA). Hanaki and Jacquemet thank both institutions for their various supports.

†GREDEG, Université Nice Sophia Antipolis, 250 Rue Albert Einstein, 06560, Valbonne, France, and Skema Business School Nobuyuki.HANAKI@unice.fr

‡Paris School of Economics and University Paris 1 Panthéon-Sorbonne, MSE, 106 Bd de l'HÃ´pital, 75013 Paris, France nicolas.jacquemet@univ-paris1.fr

§CNRS & EHESS, Aix-Marseille University (Aix-Marseille School of Economics), 2 Rue de la Charité, 13002, Marseille, France. stephane.luchini@univ-amu.fr

¶Université de Lyon and Université de Lyon 2, 69007, Lyon, France; CNRS, GATE Lyon Saint-Etienne, 93 Chemin des Mouilles, 69130, Ecully, Francezylbersztejn@gate.cnrs.fr.

1 Introduction

Coordination games provide a useful game-theoretical paradigm for analyzing a wide range of economic phenomena, such as macroeconomic fluctuations (Cooper and John, 1988), bank runs and speculative currency attacks on financial markets (Morris and Shin, 2003; Heinemann, 2012), and commercial production processes (Brandts, Cooper, and Weber, 2014). Because of the multiplicity and the Pareto-rankability of the Nash equilibria in these games, and because decisions are usually made in a state of strategic uncertainty regarding others' behavior, the resulting outcomes can be driven away from the Pareto–Nash equilibrium—a phenomenon known as coordination failure.

Coordination failure has been shown to be a persistent pattern in numerous lab implementations (Camerer, 2003, Ch.7). The present paper contributes to a large body of experimental studies exploring this welfare-reducing phenomenon. A vast part of this literature is based on the core idea that coordination failures arise from strategic uncertainty, and various institutional designs are put forth as a remedy against it: introducing repeated encounters, varying the stability and the size of groups, providing information feedback, allowing for observation of others' past behavior, or introducing pre-play communication between players (see also Devetag and Ortmann, 2007, for an extensive survey of this literature.) However, although these mechanisms are usually found to improve efficient coordination, they fall short of completely solving the problem of coordination failure.

In this paper, we take a further step to deepen our understanding of the nature of coordination failure. Our experimental results confirm that strategic uncertainty is an important determinant of the efficiency of strategic decision-making. Even more importantly, our experiment shows that individual cognitive ability has a strong link with the way strategic uncertainty influences the decisions of subjects in our experimental coordination games.

Our investigation involves a classic 2×2 coordination game, based on Selten (1975) and Rosenthal (1981), and is presented in Table 1. With $L < S < H$, $m < h$, and $s < h$, the game is one-step dominance solvable: the elimination of Player B's weakly dominated strategy l immediately leads to the Pareto–Nash equilibrium (R, r). Moreover, from the standard theory

Table 1: A simple two-person two-action dominance solvable coordination game

		Player B	
		l	r
Player A	L	(S ; s)	(S ; s)
	R	(L ; m)	(H ; h)

perspective, (R, r) is a natural candidate for a focal point, since it is also risk dominant.[1]

Notwithstanding these predictions, various studies have found a frequent failure to achieve the efficient equilibrium (see, e.g., Beard and Beil, 1994; Beard, Beil, and Mataga, 2001; Goeree and Holt, 2001; Cooper and Van Huyck, 2003; Jacquemet and Zylbersztejn, 2014) both in sequential and simultaneous implementations of this game. Depending on the exact experimental setup, between 20 and 84 % of observed outcomes are not Pareto efficient. While the literature has long focused on the strategic uncertainty faced by Player A as the source of coordination failures in these experiments, recent evidence provides different clues for explaining this behavior. For example, Polonio, Di Guida, and Coricelli (2014) use eye-tracking data gathered from simple 2×2 games to demonstrate that some subjects do not pay attention to the payoffs of their opponent, and thus do not realize that the opponent has a dominant strategy. Thus, some Player As in our coordination game may choose L without taking Player Bs' behavior into consideration at all—which precludes any meaningful role of strategic uncertainty on the decision-making of the former.

Therefore, the first aim of this paper is to understand the extent to which deviations from strategy R by Player As is due to strategic uncertainty, which constitutes an important step towards designing more efficient mechanisms aimed at eliminating coordination failure. To address this issue we conduct a set of experiments based on four variations of a dominance solvable coordination game shown in Table 1, in which human subjects (acting as Player As) interact with Player Bs represented by either (a) other human subjects, or (b) a computer program. Computerized

[1] Another Nash equilibrium, (L, l), involves a weakly dominant strategy l by Player B. The existence of a clear-cut theoretical benchmark distinguishes this game from another well-known 2×2 coordination game, the stag hunt, in which each Nash equilibrium is supported by a certain solution concept – either payoff dominance or risk dominance.

Player Bs are programmed to always choose r, and this fact is clearly explained to the subjects. Therefore, subjects acting as Player As interacting with computers do not face any strategic uncertainty, which provides an empirical benchmark for assessing the effect of strategic uncertainty on Player As' behavior in human–human interactions.

In this sense, our experiment is related to a recent and growing body of experimental studies that seek to separate and evaluate the behavioral effect of strategic uncertainty in collective decision-making. For example, in an alternating bargaining game, Johnson, Camerer, Sen, and Rymon (2002) investigated the effects of two potential causes for failure in backward induction: confusion or other-regarding social preferences. They found evidence that confusion was an important cause of deviations from the equilibrium outcome. Houser and Kurzban (2002) and Ferraro and Vossler (2010) do the same in public good contribution experiments, and estimate that confusion explained up to around one half of contribution levels. Fehr and Tyran (2001) focused on the strategic aspects of "nominal illusion". They considered four-player repeated price setting games, and introduced a negative nominal shock in the middle of the experiment. They found that roughly half of non-immediate adjustment to the new equilibrium after the shock was due to individual bounded rationality (or confusion) and the other half was due to strategic uncertainty. Finally, Akiyama, Hanaki, and Ishikawa (2015) investigated the magnitude of the effect of strategic uncertainty in explaining the observed deviation of price forecasts from the fundamental values in an experimental asset market *à la* Smith, Suchanek, and Williams (1988). They found significant effects of both confusion and strategic uncertainty.[2]

Our second objective is to shed new light on the relationship between cognitive ability and strategic thinking. In particular, we investigate whether the failure to seek efficiency by choosing R is more widespread among Player As with low cognitive ability than for those with high cognitive ability. We address this question by conducting a cognitive ability test in several experimental sessions (involving both human–human and human–robot interactions), and correlate the efficiency

[2] Other studies, somewhat less related to ours, used robots that did not follow equilibrium strategies as a way to control for subjects' beliefs about the behavior of their opponents. For instance, Ivanov, Levin, and Niederle (2010) used robots to replicate past behaviors of their subjects, and Embrey, Fréchette, and Lehrer (2014) and Costa-Gomes and Crawford (2006) used robots to make some players follow the predetermined distribution of boundedly rational behaviors.

of observed behavior and subjects' test scores, while controlling for the presence of strategic uncertainty.

From this perspective, our study contributes to recent literature that investigates the relationship between subjects' cognitive ability and their degree of strategic sophistication. For example, Brañas-Garza, García-Muñoz, and Hernán (2012) reported that subjects with higher scores on the Cognitive Reflection Test (CRT, Frederick, 2005) choose, on average, numbers closer to the Nash equilibrium in the beauty contest games. In the same vein, Akiyama, Hanaki, and Ishikawa (2015) reported that the magnitude of the effect of strategic uncertainty is positively correlated with subjects' scores on the CRT test, while the effect of confusion is negatively correlated with the score. Burks, Carpenter, Goette, and Rustichini (2009) reported that subjects (trainee truckers) with higher scores in Raven's progressive matrix test[3] are more patient and more willing to take calculated risks.[4] In addition, they reported that subjects with higher Raven's test scores more accurately predict others' behavior in a sequential prisoners' dilemma game, and better adapt their behavior to others' behavior. Carpenter, Graham, and Wolf (2013) showed that subjects with higher scores in Raven's test more frequently win in "Race to 5, 10, or 15" games[5] and guessed others' choices better in a 20-player beauty contest game. Finally, Gill and Prowse (2015) also reported that subjects with higher scores in Raven's test not only choose numbers closer to the equilibrium in a repeated 3-player beauty contest game, but also respond to the average score of other subjects in the group by choosing number close to the equilibrium when facing with others with higher scores than when facing with others with lower scores. Fehr and Huck (2015) reported

[3] Raven's progressive matrix test (often called Raven's test) is a picture based, non-verbal measure of fluid intelligence, that is "the capacity to think logically, analyze and solve novel problems, independent of background knowledge" (Mullainathan and Shafir, 2013, p. 48). It is widely used by, e.g., psychologists, educators and the military (Raven, 2000). It consists of a series of tasks to be solved within a fixed amount of time (for instance, we use a series of 16 tasks to be solved in 10 min). In each task, a subject should pick a single element (among 8 options) that best fits a set of 8 pictures. These pictures are put into a certain logical order and presented in a 3×3 table with a blank space in the bottom right corner. The level of difficulty increases from one question to the other. See Raven (2008) for an overview.

[4] Dohmen, Falk, Huffman, and Sunde (2010) reported similar correlations between cognitive ability (measured with a verbal and a nonverbal task related to the Wechsler Adult Intelligence Scale) and risk and time preferences in a representative sample of the representative German population.

[5] The "Race to 5 (or 10 or 15)" game is a two player sequential move game in which two players, moving alternatively, can put either 1, 2 or 3 stones in a common hat which is empty at the beginning. The player who puts the 5th (or 10th or 15th, respectively) stone in the hat wins. The first mover has a clear advantage in this game, and one can derive the winning strategy by a backward induction. The difficulty of deriving the winning strategy increases with the number of target stones.

similar results from a beauty contest game. They found a critical threshold of cognitive ability (measured by CRT) below which subjects choose random numbers and do not respond to their beliefs about others' cognitive ability. Subjects with cognitive ability above this threshold, however, tend to act much more strategically: they systematically choose lower numbers and respond to their beliefs about the cognitive ability of other players. Finally, recent evidence from psychological research reveals the relationship between fluid intelligence and the theory of mind (Ibanez, Huepe, Gempp, Gutiérrez, Rivera-Rei, and Toledo, 2013).[6]

To sum up, these empirical studies suggest that people with high cognitive ability respond more aptly to strategic conditions they face than those with low cognitive ability. The present study extends this investigation to a new and important economic environment—the coordination game. As will be seen, we find that Player As' failure to choose R can be only partially explained by uncertainty about their partners' intentions: in many cases, the former act in this manner even when interacting with a computer program that is known to always act reliably by choosing r. We also report that Player As with high cognitive ability (measured in terms of Raven's test scores) tend to be more sensitive to strategic uncertainty than those with low cognitive ability.

2 Experimental design

We consider four payoff matrices based on the simultaneous-move coordination game shown in Table 1. Our main manipulation lies in varying the nature of Player B, who may be represented either by a human subject (Human treatment) or a pre-programmed computer (Robot treatment). All games and treatments are implemented using a between-subject design—only one version of the game is played in each experimental session. In all the sessions, the one-shot game is repeated ten times with participants' roles remaining fixed, pairs being rematched in each round using a

[6]Baron-Cohen, Wheelwright, Hill, Raste, and Plumb (2001) developed the "Reading the Mind in the Eyes" test (RMET) to measure one's theory of mind—the capacity to infer the internal emotional states of others. RMET consists of a series of photos of the area of the face involving the eyes. Subjects are asked to choose one of the four words that best describes what the person in the photo is thinking or feeling. Ibanez, Huepe, Gempp, Gutiérrez, Rivera-Rei, and Toledo (2013) found that people with high scores in Raven's test also perform better in RMET. In an experimental investigation of the Level-k model, Georganas, Healy, and Weber (2015) found a positive correlation between the score in RMET test and the propensity to adapt Level-1 reasoning.

Table 2: The experimental games

Baseline 1

		B	
		l	r
A	L	(9.75 ; 3.00)	(9.75 ; 3.00)
	R	(3.00 ; 4.75)	(10.00 ; 5.00)

Baseline 2

		B	
		l	r
A	L	(8.50 ; 3.00)	(8.50 ; 3.00)
	R	(6.50 ; 4.75)	(10.00 ; 5.00)

Egalitarian 1

		B	
		l	r
A	L	(9.75 ; 8.50)	(9.75 ; 8.50)
	R	(3.00 ; 8.50)	(10.00 ; 10.00)

Egalitarian 2

		B	
		l	r
A	L	(8.50 ; 8.50)	(8.50 ; 8.50)
	R	(6.50 ; 8.50)	(10.00 ; 10.00)

perfect stranger, round-robin procedure[7] and take-home earnings corresponding to a single round randomly drawn at the end of each experimental session. At the end of each round, subjects are only informed of their own payoffs.

2.1 Treatments and hypotheses

The four experimental game matrices are presented in Table 2. Two of them have already been experimentally studied in the literature. Baseline Game 1 (BG1, shown on the top left panel of Table 2) was used as the baseline treatment in Beard and Beil (1994); Beard, Beil, and Mataga (2001), Jacquemet and Zylbersztejn (2013, 2014). Egalitarian Game 2 (EG2, shown on the bottom right panel of Table 2) was one of the additional matrices introduced by Jacquemet and Zylbersztejn (2014) in an attempt to assess the effect of the relative payoff structure on subjects' behavior. The latter study reports a strong divergence between both players' behavior in these two games: weak reliability from Player Bs (80.7 % of decisions r) coupled with weak reliance from

[7] Repetition allows to assess the extent to which inefficient behavior is sensitive to learning. For this sake, we use an indefinitely repeated game with one-round compensation rule1 as an attempt to homogenize incentives across rounds, and allow for an accumulation of experience from a series of uniform one-shot interactions. Kamecke (1997) shows that our perfect stranger, round-robin procedure is optimal for this purpose since it maximizes the number of rounds for a given number of players and the one-shot nature of each interaction between subjects.

Player As (49 % of decisions R) in BG1, and nearly universal reliability from Player Bs (94.3 % of r) coupled with strong (yet imperfect) reliance from Player As (74 % of R) in EG2.[8] Importantly, Jacquemet and Zylbersztejn (2014) provide systematic evidence that these outcomes cannot be explained by inequality aversion.

The payoff structures of BG1 and EG2 differ in terms of both players' monetary incentives to seek efficiency. In BG1, Player As may improve their situation only slightly when moving from L to (R, r) (from 9.75 to 10, a difference of .25), while a failed attempt to rely on Player Bs, resulting in (R, l), is very costly (yielding only 3 to Player As). In EG2, these two cases become more balanced—the gain for moving from L to (R, r) increases (from 8.5 to 10, a difference of 1.5), and the cost of relying on the other player in vain becomes less severe, with (R, l) now giving 6.5 to Player As. Analogous variations occur for Player Bs. The efficiency premium (conditional on Player As' reliance) is quite low in BG1 (from 4.75 to 5, a difference of .25) and drastically rises in EG2 (from 8.5 to 10, a difference of 1.5). Altogether, EG2 provides much more salient monetary incentives to act efficiently to both players.[9] To account for subjects' responses to the changes in their own as well as their partners' monetary incentives, we introduce two intermediate payoff matrices: Baseline Game 2 (BG2) and Egalitarian Game 1 (EG1) shown, respectively, in the top-right and bottom-left corners of Table 2. Each of the two games differs in only one dimension—that is, either Player As' payoffs or Player Bs' payoffs—as compared to BG1 and EG2: BG2 (EG1) has Player A's payoffs taken from EG2 (BG1) and Player B's payoffs taken from BG1 (EG2). Thus, the four games enable us to test the following hypotheses for the Human condition:

Hypothesis 1 *In the Human treatment, the variations in the monetary incentives affect players' behavior as follows:*

(a) Player As react to the variations in their own monetary incentives to seek efficiency: the proportion of decisions R is higher in BG2 than in BG1, and in EG2 than in EG1;

[8] In the present study, we focus on the determinants on Player As' behavior, considering Player Bs' solely as a source of strategic uncertainty. Jacquemet and Zylbersztejn (2013) offer a systematic analysis of the patterns of Player Bs' decisions in this game.

[9] Some studies have also documented the effect of the saliency of monetary incentives in coordination games. See, for example, Battalio, Samuelson, and Van Huyck (2001) for symmetric 2×2 games, and Goeree and Holt (2005); Devetag and Ortmann (2010) for n-player minimum and median effort games.

(b) *Player Bs react to the variations in their own monetary incentives to act efficiently: the proportion of decisions* r *is higher in EG1 than in BG1, and in EG2 than in BG2;*

(c) *Player As react to the variations in Player Bs' monetary incentives to act efficiently: the proportion of decisions* R *is higher in EG1 than in BG1, and in EG2 than in BG2.*

The strategic uncertainty that Player As face cannot be directly observed or measured by the experimenter. Therefore, assessing its behavioral effect requires a benchmark in which the actual degree of strategic uncertainty can be controlled for. To that end, each of the four games is implemented under two different conditions: Human and Robot. In the Human treatment, two human subjects interact in ways described above. In the Robot treatment, a human subject acting as Player A interacts with a computerized Player B who is pre-programmed to always choose r. Subjects in the Robot treatment are clearly informed they are interacting with a pre-programmed computer: "**the computer chooses** r **at each round, without exception**" (bold in the original instruction sheet). This is the only difference in the rules and procedures between Human and Robot treatments. As a result, subjects in the Robot treatment do not face any strategic uncertainty. This leads us to the following hypothesis:

Hypothesis 2 *The Robot treatment neutralizes strategic uncertainty Player As face. As a result, the proportion of actions* R *in a given game is higher in the Robot treatments than in the Human treatments.*

While not choosing R in the Human treatment *may* not necessarily arise from strategic uncertainty, the same behavior in the Robot treatment *must* be due to reasons other than strategic uncertainty. Thus, comparing Player As' decisions between Human and Robot treatments enables us to capture the behavioral effect of strategic uncertainty.

Once the behavioral results are established, we then investigate the relationship between subjects' cognitive ability and their sensitivity to the changes in the degree of strategic uncertainty. We measure each participant's cognitive ability by implementing Raven's test at the end of all experimental sessions involving the BG2 and EG1 games, under both Robot and Human treatments. Based on existing experimental results (discussed in the opening section), we formulate the following hypothesis:

Hypothesis 3 *Subjects with high cognitive ability are more sensitive to changes in the degree of strategic uncertainty than those with low cognitive ability.*

2.2 Experimental procedures

Upon arrival, participants are randomly assigned to their computers and asked to fill in a short personal questionnaire containing basic questions about their age, gender, education, etc.[10] The pre-distributed written set of instructions is then read aloud. Player As are informed that they will play an unrevealed number of rounds of the same game, each round with a different partner, and that their own role will not change during the experiment. Before starting, subjects are asked to answer a quiz assessing their understanding of the game they are about to play. Once the quiz and any questions from participants are answered, the experiment begins.

The experiment generates observations under eight experimental conditions, varying according to the payoff structure (BG1, BG2, EG1 or EG2) and Player A's partner (human subject in the Human treatment or computer in the Robot treatment). All conditions are implemented separately, using a between-subject design: each subject plays only one of the four games, and interacts either with other subjects or with a computer.

For each payoff matrix, we ran three Human treatment sessions (involving 20 subjects per session: 10 Player As interacting with 10 Player Bs), and two Robot treatment sessions (involving 20 Player As per session interacting with automated Player Bs). The data for the BG1 and EG2 Human treatments come from Jacquemet and Zylbersztejn (2014), while all the other sessions were carried out in October 2012 and February and March 2014.[11] Of the 398 participants (190

[10]We decided to implement the administrative questionnaire at the beginning of the experiment to reduce the noise in answers and to avoid an accumulation of post-experimental surveys. As correctly stressed by a referee, this might raise concerns about anonymity in subjects' decision-making. However, this part of the design is identical in all session and thus should not affect our main results that are based on the between-treatment differences.

[11]The unexpected behavior initially observed for matrices BG1 and EG2 led us to complement our design with matrices BG2 and EG1, hence the delay between the two sets of experiments. To assure an in-depth exploration of players' behavior, these complementary sessions also included Raven's test.

males), 323 were students with various fields of specialization.[12] The majority of subjects (57 %) had already taken part in economic experiments. Participants' average age was 24.05.[13] Sessions lasted about 45–60 minutes, with an average payoff of roughly 12.50 euros in Human treatments and 15 euros in Robot treatments (including a 5 euros show-up fee, but not the post-experiment task fee).[14] No subject participated in more than one experimental session.

2.3 Control variable on cognitive ability: Raven's test score

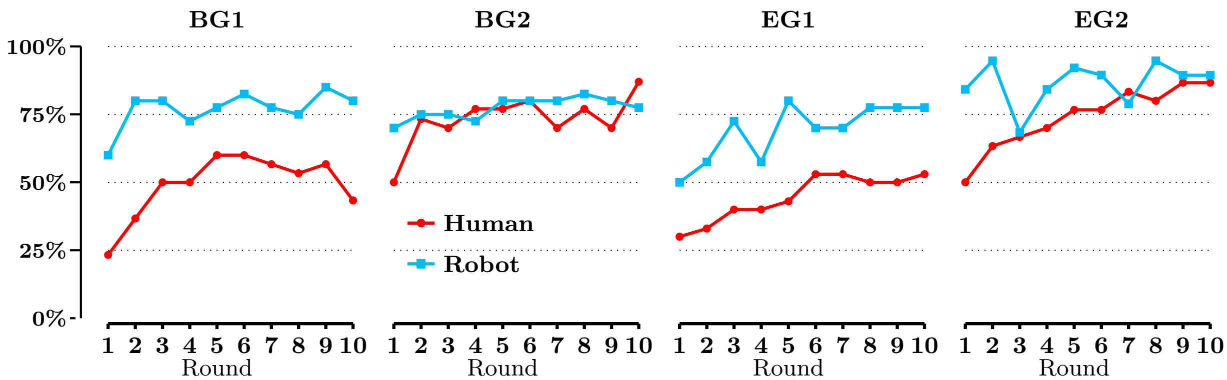
All sessions involving BG2 and EG1 matrices (both under Human and Robot treatments) include computerized post-experiment tasks. An additional 5 euros is paid to each subject for completing this part. Immediately after the end of the experimental game, participants are provided with a brief round-by-round summary of their decisions and outcomes, and are asked to provide any relevant comments and indicate the things that might have affected their decisions during the experiment in a blank space on their computer screens. Subjects are also asked to solve a part of the advanced version of Raven's test—composed of 16 items to be completed within 10 minutes. Overall, the data from Raven's test include 180 subjects: 120 (60 player As and 60 players Bs) in both Human treatments, and 80 (all player As) in both Robot treatment. As we argue in the next section of the paper, cognitive ability (measured by Raven's test score) comes as a granular explanation of aggregate inefficiencies under different forms of strategic uncertainty and monetary incentives. However, a necessary condition for this argument to hold is that the distributions of these scores should not vary across game matrices and Human/Robot treatment. We find strong evidence that the distribution of cognitive ability does not vary across experimental conditions. A multiple-treatment comparison using Kruskal–Wallis test with Bonferroni correction does not

[12]In one EG2 Robot treatment session, we had 18 subjects instead of 20, so for the Robot treatment there were 40 subjects for BG1, BG2 and EG1, and 38 subjects for EG2. For the Human treatment sessions, we had 60 subjects (half of whom are Player As) for each of the four games.

[13]All sessions took place at the *Laboratoire d'Economie Experimentale de Paris* (LEEP) at Paris School of Economics. Subjects were recruited via an online registration system based on ORSEE (Greiner, 2004) and the experiment was computerized through software developed under REGATE (Zeiliger, 2000) and z-Tree (Fischbacher, 2007).

[14]As will be described below, Raven's test was included in half of our experimental sessions and was carried out as a post-experimental task. For this post-experimental task, 15 additional minutes were needed beyond the usual duration of the sessions (around 45 min, including the time to read the instructions, answer the questionnaires, play 10 rounds of the experimental game and be paid for participation).

Figure 1: Proportion of decisions R across rounds and treatments



reject the null hypothesis that Raven's test scores in the four experimental conditions come from the same population (with $p = .275$). The same test applied at the session-level (10 sessions) instead of the experimental condition-level yields a $p = .694$.

3 Results

Figure 1 provides descriptive statistics regarding the behavior of Player As in all our experimental treatments. The share of Player As who chose R in each round is displayed separately for each payoff matrix, and the two curves provide a comparison between the Human treatment and the Robot treatment. Before moving to a detailed analysis of the treatment effects, three main observations can be made. First, our Human treatment replicates the results seen in the existing literature: a high proportion of Player As decide to play L, even after several rounds of the game. Second, between game comparisons of behavior in the Human treatments show this pattern is barely influenced by the strategic context: while Player As react to changes in incentives they face (BG2 vs BG1, and EG2 vs EG1), they appear rather insensitive to changes in incentives faced by Player Bs (EG1 vs BG1, and EG2 vs BG2). Finally, while the share of decisions R in the Robot treatment always weakly dominates the one in the Human treatment, the absence of strategic uncertainty in this context does not remove all decisions L. Table 4 in the "Appendix" provides a robustness check of these effects based on parametric probit models estimated separately for each game on the pooled Human–Robot data. The results suggest that the likelihood of action R in the initial round of each game is higher in the Robot treatment (dummy variable *Robot* is significant for each game). However, the subsequent dynamics do not differ between the two conditions: the Wald test rejects the joint insignificance of *Robot dummy* $\times$ *round effects* dummies solely for the EG2 data.[15]

In the remainder of this section, the main question we seek to answer is whether and to what extent strategic uncertainty drives the observed decisions to play L. In the Human treatment, two factors explain the decisions of Player As: the behavior of Player Bs in the experiment and how Player As adjust to this behavior. The variations in payoffs between games allows these two factors to be measured separately. We then move to an analysis of the Robot treatments, in which strategic uncertainty is removed by design.

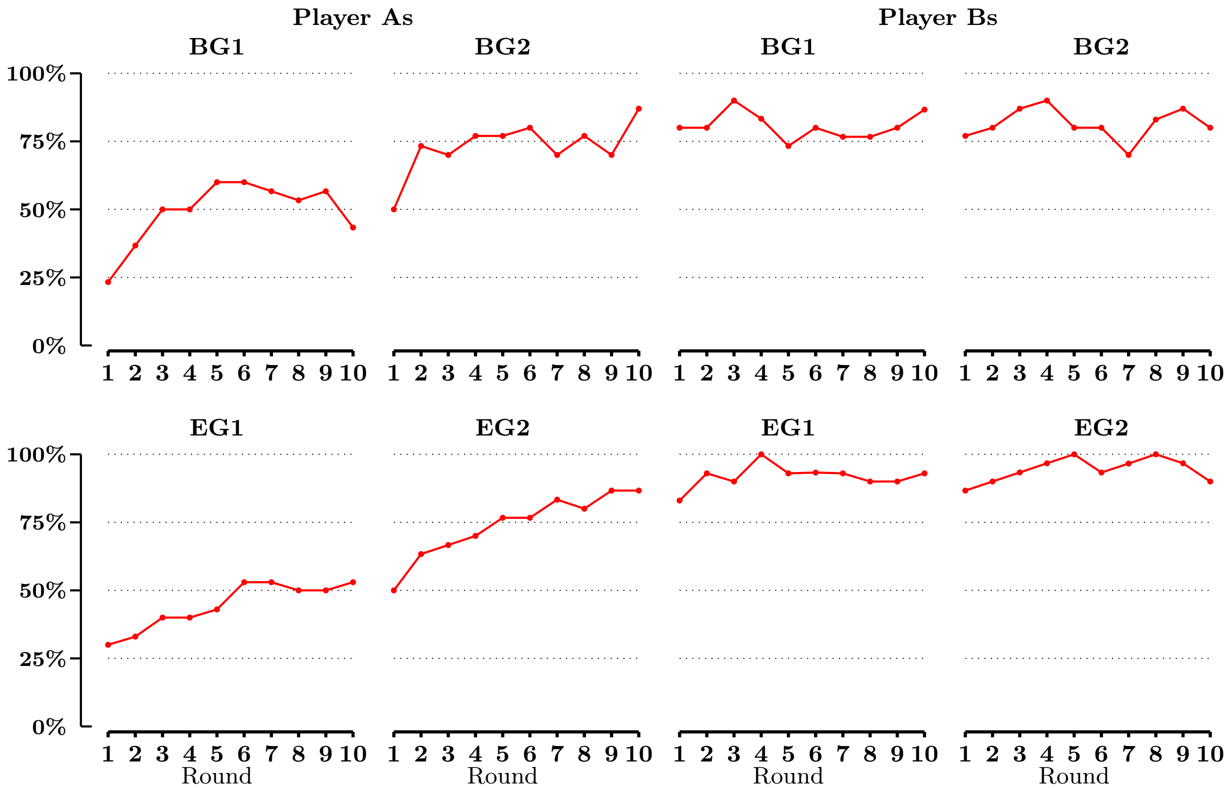
3.1 Results from the Human treatment

As summarized in Hypothesis 1, the variations in payoffs between games should induce variations in the decisions of both players, hence resulting in a variation in the actual strategic uncertainty faced by Player As. Figure 2 provides an overview of individual behavior in the Human treatment. Aggregate results suggest that both players react to the variations in their own payoff scheme. Holding Player Bs' payoffs constant, Player As are more likely to seek efficiency as their monetary incentives to do so become more salient: the frequency of R increases from 49 % in BG1 to 73 % in BG2 ($p = .001$), and from 45 % in EG1 to 74 % in EG2 ($p = .005$).[16] Analogously, Player Bs become more efficient the higher the cost of acting otherwise: the frequency of r increases from 81 % in BG1 to 92 % in EG1 ($p = .026$) and from 81 % in BG2 to 94 % in EG2 ($p = .012$). However, despite Player Bs' responsiveness to their personal monetary incentives, Player As remain insensitive to this factor: the differences between BG1 and EG1 and between BG2 and EG2 are

[15] BG1: $p = .402$; BG2: $p = .385$; EG1: $p = .557$; EG2: $p = .002$.

[16] We test the difference in proportion of a given outcome between two experimental conditions by carrying out a two-sided bootstrap proportion test that accounts for within-subject correlation—*i.e.* the fact that the same individual takes 10 decisions. The procedure consists of bootstrapping subjects and their corresponding decisions over all ten rounds instead of bootstrapping decisions as independent observations (see, e.g., Jacquemet, Joule, Luchini, and Shogren, 2013, for a detailed description of the procedure). In Round 1, data are independent and thus allow us to analyze behavior with a standard bootstrap proportion test. Frequencies in Round 1 are 23.3 % in BG1 and 50.0 % in BG2 ($p = .027$), and 30.0 % in EG1 and 50.0 % in EG2 ($p = .091$).

Figure 2: Share of the decisions R (r) for Player As (Bs) in the Human treatments, across rounds



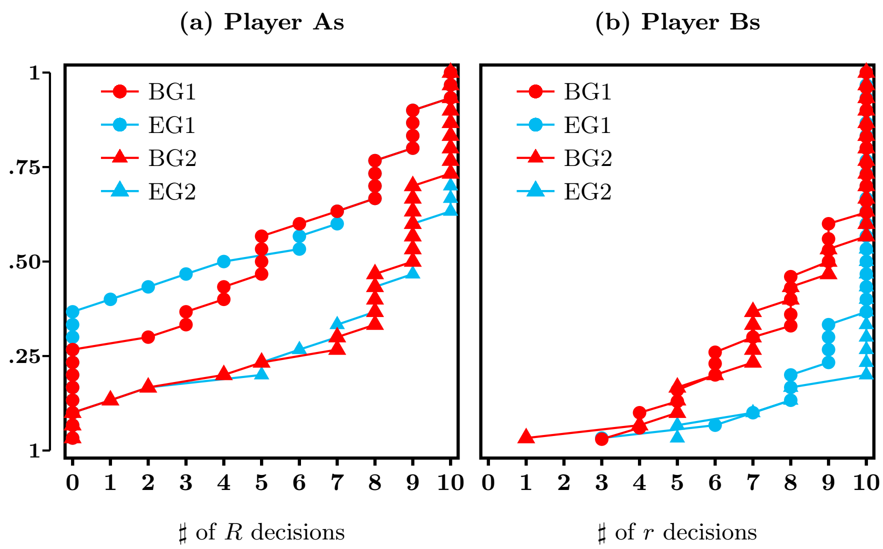
small and not statistically significant ($p = .694$ and $p = .898$, respectively).[17]

These three observations are supported by the individual-level data summarized in Fig. 3. The left-hand side provides the Empirical Distribution Function (EDF) of Player As' frequency of decisions R, i.e. the number of decisions R taken throughout the ten rounds of the game. Statistical tests indicate first-order stochastic dominance of the distributions between BG2 and BG1 ($p = .007$) and between EG2 and EG1 ($p = .021$).[18] This confirms that Player As' behavior depends on the saliency of their own monetary incentives. Figure 3b presents the EDFs of Player Bs' individual frequencies of decisions r. These individual-level data stand in line with the

[17]The analogous frequencies in Round 1 for Player Bs are 80.0 % in BG1 and 83.3 % in EG1 ($p = .731$), and 76.7 % in BG2 and 86.7 % in EG2 ($p = .232$); for Player As, they are: 23.3 % in BG1 and 30.0 % in EG1 ($p = .583$), and 50.0 % in BG2 and 50.0 % in EG2 ($p = .889$).

[18]Our first-order stochastic dominance test is based on a bootstrap version of the univariate Kolmogorov–Smirnov (KS) test which allows for ties (see, *e.g.*, Abadie, 2002; Sekhon, 2011).

Figure 3: EDF of the total number of decisions R and r decisions by game



aggregate outcomes, showing that Player Bs do react to their personal costs of inefficient behavior: the EDF from EG1 first order stochastically dominates the EDF from BG1 (p = .039), and the EDF from EG2 first order stochastically dominates the EDF from BG2 (p = .002). This amounts to a significant change in the actual strategic uncertainty Player A faces. Figure 3a also shows that there are no significant differences between EG1 and BG1 (p = .783) or between EG2 and BG2 (p = .914); Player As do not seem to react to the changes in the saliency of Player Bs' monetary incentives.

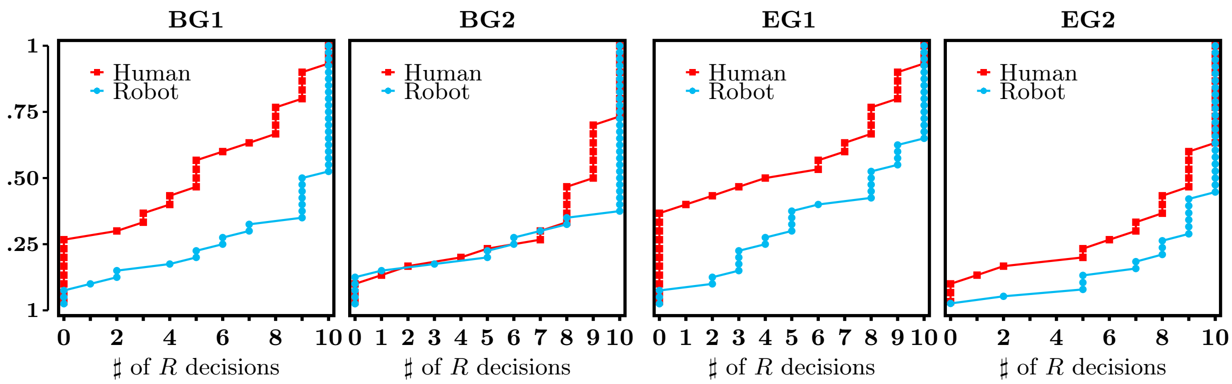
Result 1 *In the presence of strategic uncertainty, the saliency of monetary incentives affects players' behavior as follows:*

(a) The proportion of decisions R is higher in BG2 than in BG1, and in EG2 than in EG1: Player As become more likely to act efficiently the more salient their monetary incentives;

(b) The proportion of decisions r is higher in EG1 than in BG1, and in EG2 than in BG2: Player Bs become more likely to act efficiently the more salient their monetary incentives;

(c) The proportion of decisions R is not significantly different in EG1 than in BG1, nor in EG2

Figure 4: EDF of the individual number of decisions R by treatment



than in BG2: Player As ignore the saliency of Player Bs' monetary incentives.

3.2 The effect of strategic uncertainty: comparing Human and Robot treatments

In the Robot treatment, Player Bs' decisions are generated by computers always playing r, hence eliminating the strategic uncertainty faced by Player As. The change in behavior from Player As hence identifies the effect of strategic uncertainty in the Human treatment. In Fig. 1 we presented aggregate statistics on the behavior of Player As. Figure 4 provides individual data for all four games and both treatments. Each curve provides the EDF of the number of decisions R taken by each individual in the Human treatment and in the Robot treatment. In the aggregate, in three games out of four we find a statistically significant increase in the proportion of decisions R when Player As interact with computerized Player Bs rather than other human subjects: from 49.0 to 77.0 % in BG1, from 44.7 to 69.0 % in EG1 and from 77.0 to 86.6 % in EG2. In BG2, however, we only observe a weak increase—from 72.5 to 77.3 %.[19].

These results are true all along the distribution of the number of decisions R. The EDF in the Robot treatment first order stochastically dominates the EDF in the Human treatment for BG1 ($p < .001$), BG2 ($p = .004$) and EG1 ($p = .011$); this pattern is not statistically significant for

[19]These shifts are significant according to bootstrap proportion tests at the 1 % threshold in BG1 ($p = .001$) and in EG1 ($p = .005$), at the 5 % threshold in EG2 ($p = .033$) and are not statistically significant in BG2 ($p = .305$)

Table 3: Spearman's rank correlation of the number of decisions R and Raven's test scores

BG2	Human	.055 ($p = .772$, $N = 30$)
	Robot	.571 ($p < .001$, $N = 40$)
EG1	Human	-.005 ($p = .978$, $N = 30$)
	Robot	.273 ($p = .088$, $N = 40$)

EG2 ($p = .134$).[20] The most noticeable difference between the two distributions is the proportion of Player As who invariably choose R when facing a computer rather than another subject: 50.0 % against 10.0 % in BG1, 65.5 % against 30.0 % in BG2, 37.5 % against 10.0 % in EG1 and 57.9 % against 40.0 % in EG2. This increase is significant at a 1 % threshold for games BG1, BG2 and EG1 ($p = .001$, $p = .004$ and $p = .010$, respectively) and barely insignificant in EG2 ($p = .111$).

Importantly, both aggregate and individual results unambiguously show that eliminating strategic uncertainty does not make Player As' behavior invariably efficient. Altogether, these observations lead us to the following result:

Result 2 *Strategic uncertainty only partially explains Player As' failure to seek efficiency in the coordination game. Player As do not consistently select the Pareto–Nash strategy R even when interacting with a computer who is known to always act efficiently by choosing r.*

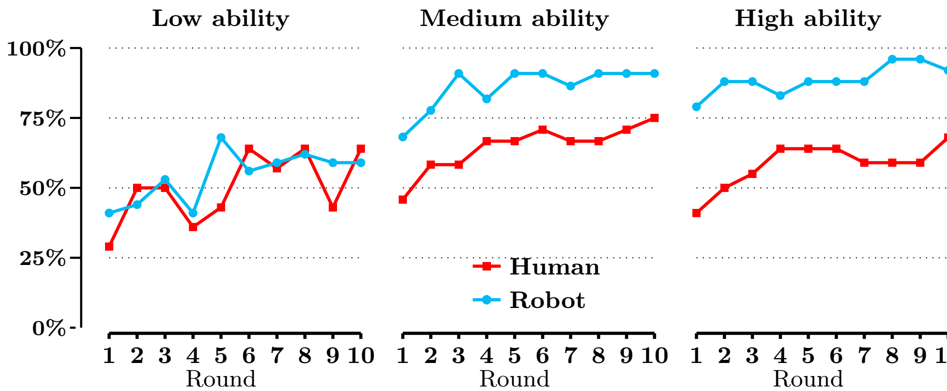
3.3 Cognitive ability and strategic uncertainty

The crucial question raised by the Human–Robot comparison is why do Player As continue acting inefficiently in the absence of strategic uncertainty? We investigate this question based on the cognitive ability of Player As. Cognitive ability is measured by Raven's test, performed after the game only in BG2 and EG1, in both Human and Robot treatments. It corresponds to the number of correct answers to the set of 16 questions.

Table 3 reports the individual correlations between the scores in Raven's test and decisions in the game, measured by the number of choices R (between 0 and 10) made throughout the 10

[20]First-order stochastic dominance (FOD) in BG2 is induced by a sharp increase in the proportion of subjects playing R in all 10 rounds (from 30.0 % in the Human treatment to 65.5 % in the Robot treatment). This is enough to induce a statistically significant FOD. In EG2, the magnitude of this increase is small: from 40 % of subjects playing R in all 10 rounds in the Human treatment to 57.9 % in the Robot treatment.

Figure 5: Proportion of decisions R across rounds and treatments by ability group

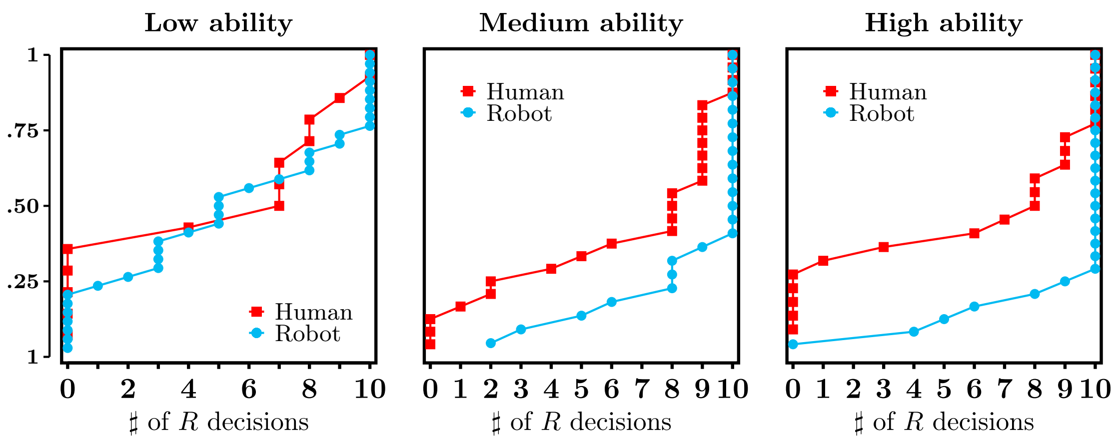


rounds of the game. We find a positive and significant correlation between these two variables in the Robot treatments which suggests that subjects with higher cognitive ability are more likely to play R. However, this correlation disappears in Human treatments. To gain further insights on the transition between the Robot and Human treatments, we categorize individuals in our experimental sample according to their score on the Raven's test. Player As are divided into three groups, which correspond to the three tertiles of the overall score distribution, (i.e. considering all Player As from all four experimental conditions, hence 140 subjects.) Subjects in the low-ability group have a Raven's test score below 7, those with a Raven's test score of 7–10 are in the medium-ability group and those with a score above 10 are in the high-ability group.

Figure 5 presents the proportion of decisions R across rounds by cognitive ability group.[21] The aggregate dynamics are very similar across cognitive ability groups, with an increase in the frequency of decisions R in the first rounds and a stabilization afterwards. This result holds regardless of whether decisions are taken with computers or humans acting as Player Bs. However, interacting with computers instead of humans induces an initial upward shift in the ratio of R that persists over time in the medium- and the high-ability groups, whereas no such shift occurs in the low-ability group. The mean increase in proportion of decisions R in Round 1 induced by playing with a computer rather than a human is 22.4 % (p = .057) for the medium-ability

[21] The data from BG2 and EG1 are pooled to focus on the overall effect of removing strategic uncertainty and guarantee sufficient sample sizes in each category.

Figure 6: EDF of the total number of decisions R by treatment and cognitive ability group



group and 38.3 % for the high-ability group (p = .003). The mean difference between these proportions is still present in Round 10, but to a lesser extent: 15.9 % ($p = .074$) for the medium-ability group and 11.5 % (p = .022) for the high-ability group. In the low-ability group, we find a small and insignificant upward shift in the proportion of decisions R in Round 1 (12.6 %, p = .203), while the proportions of decisions R are almost equally likely in Round 10 (the difference being -5.5 %, p = .618) in Robot and Human treatments. We assess the robustness of these patterns by estimating parametric probit models separately for each of the cognitive ability groups using the pooled Human–Robot data from both games. The results are reported in Table 5 in the "Appendix". First, we confirm the relationship between the behavior observed in Human and Robot treatments in each ability group (the *Robot* dummy is insignificant in the low-ability group model, and positive and significant in the two remaining models). Second, we report that the dynamics of behavior do not differ between Human and Robot treatments across different cognitive ability groups. In particular, a Wald test does not reject the joint insignificance of *Robot dummy* × *round effects* dummies for low- (p = .316), medium- (p = .906) and high- (p = .589) ability groups.

Figure 6 provides the EDFs of the number of decisions R made by each subject across ten rounds for each cognitive ability group in Human and Robot treatments. Three results emerge. First, the EDFs in the Human treatment are not statistically different across cognitive ability groups.[22] Second, this is not the case in the Robot treatment: the EDF for the low cognitive ability group is first order stochastically dominated by the EDF for the medium and high cognitive ability groups (p = .003 and p = .001, respectively). Finally, we observe no such relationship for the medium- and high-ability groups: the EDFs are not significantly different (p = .318). The main reason for these results is that the subjects in the low-ability group do not respond to the absence of strategic uncertainty in Robot treatments by increasing the frequency of decisions R (p = .415), whereas subjects in the medium- and high-ability groups do so (the tests are based on comparisons between treatments for each cognitive ability group, the p-values for both are p = .001.)

Such behavior is in line with Hypothesis 3, i.e. Player As with higher cognitive ability are more sensitive to the change in the degree of strategic uncertainty between Human and Robot treatments. To provide a formal test of the hypothesis, we first denote $\mathcal{R}$ the mean number of R decisions across 10 rounds by treatment (Robot vs. Human) and type (low cognitive ability vs. high cognitive ability). The test of Hypothesis 3 is then defined as follows:

$$\begin{cases} H_0 : \mathcal{R}(\text{Robot, high}) - \mathcal{R}(\text{Human, high}) = \mathcal{R}(\text{Robot, low}) - \mathcal{R}(\text{Human, low}) \\ H_1 : \mathcal{R}(\text{Robot, high}) - \mathcal{R}(\text{Human, high}) > \mathcal{R}(\text{Robot, low}) - \mathcal{R}(\text{Human, low}) \end{cases} \tag{1}$$

Empirical results indicate that our hypothesis is likely to be verified with $\mathcal{R}$(Robot, high) = 8.8, $\mathcal{R}$(Human, high) = 5.8, $\mathcal{R}$(Robot, low) = 5.4, and $\mathcal{R}$(Human, low) = 5.0, which is consistent with the hypothesis. We statistically tested the result by putting all the terms of Eq. (1) to the left-hand side, so that the test reduces to a comparison test of multiple means with coefficients $(1; -1; -1; 1)$.[23] The test rejects the null hypothesis when the low cognitive ability group is compared to the high cognitive ability group (p = .051), as well as when it is compared to the

[22] To ensure a sufficient sample size in each ability group, we pooled the outcomes from both games in each treatment. The tests are performed using two-sided bootstrap K-S tests. The p-values of the two-by-two comparisons are: p = .288 for the low-ability group versus the medium-ability group, p = .599 for the low versus the high and p = .695 for the medium versus the high.

[23] Given the group sizes, the bootstrap test is based on re-sampling subjects and the number of times they choose decision R. To account for asymmetry in the empirical distribution, we computed an equal-tail bootstrap p-value. See Davidson and MacKinnon (2006) for further details on this procedure.

medium and high cognitive ability groups pooled together ($p = .056$). These results allow us to state our last result:

Result 3 *Player As with high cognitive ability are more sensitive to the changes in strategic uncertainty than those with low cognitive ability.*

4 Conclusion

Coordination failure is a widely documented phenomenon, with possibly dramatic economic consequences. While most experimental investigations try to identify the possible sources of strategic uncertainty underlying such failures, very few explore the effect of strategic uncertainty *per se*. In addition, little is known about whether and how behavior under strategic uncertainty is related to individual cognitive ability.

We used four variations of a classic 2×2 dominance solvable coordination game to explore the link between coordination failure, strategic uncertainty and cognitive ability. To isolate the behavioral effect of strategic uncertainty, we compare the decisions made when facing other human subjects (Human treatment) to those made against computer programs whose perfectly efficient behavior is common knowledge (Robot treatment). We find that the occurrence of coordination failure cannot be entirely explained by strategic uncertainty. First, a non-negligible share of our subjects failed to act efficiently even in the Robot treatment. Second, behavioral response to strategic uncertainty is related to individual cognitive ability. Subjects with higher cognitive ability (as measured by their scores on Raven's test) systematically adapt their decisions to the varying degree of strategic uncertainty (i.e., between the Human and Robot treatment), while those with lower cognitive ability fail to do so.

The relationship between cognitive ability and the sensitivity to strategic uncertainty is a challenge for institutions aimed at fostering efficient coordination by reducing strategic uncertainty. These mechanisms—such as the widely studied effect of communication between participants—usually involve focusing agents' actions on the desirable equilibrium. However, the presence of players with a low strategic focus—such as the Raven's test low scorers in our experiment—may undermine the effect and lower the economic value of such institutions. As noted by Ellingsen and Östling (2010) in the context of a theoretical analysis of communication and coordination in a level-k setup, "there is a need for evidence that systematically distinguishes the effects of preferences, belief, and rationality" (p. 1714). We believe that our experimental design and collected data on the cognitive underpinnings of behavior under strategic uncertainty and the resulting coordination failure is an important step in this direction.

Appendix: Additional probit estimates

Table 4: Probit models on the decisions to play R, by game

	BG1 ($n = 700$)		**BG2** ($n = 700$)		**EG1** ($n = 700$)		**EG2** ($n = 680$)	
Variables	Coef.	p-value	Coef.	p-value	Coef.	p-value	Coef.	p-value
Round effects (round 1 is referent)								
Round 2	.787	.089	1.264	.008	.104	.821	.642	.173
Round 3	1.419	.003	1.070	.023	.515	.273	.863	.076
Round 4	1.448	.003	1.560	.002	.505	.270	1.102	.028
Round 5	1.954	.000	1.586	.002	.712	.123	1.636	.003
Round 6	1.965	.000	1.752	.001	1.411	.005	1.517	.003
Round 7	1.795	.000	1.029	.027	1.411	.005	2.331	.000
Round 8	1.599	.001	1.465	.002	1.136	.018	1.904	.001
Round 9	1.760	.000	1.072	.020	1.180	.015	2.765	.000
Round 10	1.058	.023	2.697	.000	1.397	.004	2.765	.000
Robot effect (dummy—Human treatments are referent)								
Robot	2.136	.002	1.713	.032	1.512	.028	2.152	0.003
Robot dummy × round effects								
__round 2	.728	.273	-.689	.340	.285	.640	.702	0.376
__round 3	.044	.948	-.563	.437	.754	.226	-1.920	0.005
__round 4	-.561	.397	-1.324	.070	-.117	.847	-1.053	0.139
__round 5	-.683	.315	-.509	.513	1.077	.090	-.696	0.378
__round 6	-.227	.748	-.676	.383	-.253	.696	-.866	0.255
__round 7	-.506	.454	.024	.973	-.291	.654	-2.693	0.001
__round 8	-.642	.324	-.115	.878	.480	.455	-.383	0.671
__round 9	.244	.733	.004	.995	.437	.499	-2.182	0.012
__round 10	.448	.505	-1.919	.026	.219	.735	-2.113	0.016
Constant	-1.602	.003	.067	.912	1.384	.008	.039	.939
Joint nullity Wald tests of *Robot dummy × round effects*								
Wald test	9.39	.402	9.58	.385	7.77	.557	25.44	.002
σ_{RE} (sd.)	2.074 (.310)		2.740 (.445)		2.101 (.306)		2.146 (.358)	
ρ (sd.)	.811 (.045)		.882 (.034)		.821 (.049)		.815 (.044)	

Note. Probit models on the probability of playing R estimated separately for each payoff configuration, on pooled Human and Robot treatments data. (Exogeneous) unobserved individual heterogeneity is accounted for through random individual effects. The covariates are: round fixed effects, Robot treatment dummy and interactions between the two.

Table 5: Probit models on the decisions to play R, by Raven ability groups

	Raven cognitive ability group					
	Low ($n = 480$)		**Medium** ($n = 460$)		**High** ($n = 460$)	
Variables	Coef.	p-value	Coef.	p-value	Coef.	p-value
Round effects (round 1 is referent)						
round 2	1.130	.109	.504	.272	.600	.316
round 3	1.168	.103	.511	.270	.953	.124
round 4	.306	.651	.939	.049	1.793	.011
round 5	.701	.303	.963	.048	1.729	.010
round 6	2.517	.007	1.116	.021	1.793	.011
round 7	1.757	.026	.929	.051	1.224	.050
round 8	2.517	.007	.919	.050	1.253	.040
round 9	.718	.297	1.130	.018	1.340	.040
round 10	2.517	.007	1.475	.004	2.244	.002
Game effect (dummy - BG2 is referent)						
EG1	-.118	.882	-1.948	.000	-1.821	.068
Robot effect (dummy - Human treatments are referent)						
Robot	1.184	.250	1.231	.064	3.804	.002
Robot dummy × round effects						
__round 2	-.881	.290	.016	.983	.531	.606
__round 3	-.466	.579	1.214	.143	.178	.864
__round 4	-.308	.704	-.113	.882	-1.227	.236
__round 5	.912	.273	.752	.375	-.597	.576
__round 6	-1.620	.116	.584	.483	-.661	.544
__round 7	-.701	.440	.332	.681	-.142	.887
__round 8	-1.304	.206	.814	.338	1.945	.205
__round 9	.342	.679	.585	.486	1.858	.231
_round 10	-1.454	.159	.259	.767	-.470	.677
Constant	-1.609	.087	.721	.129	.007	.995
Joint nullity Wald tests of *Robot dummy × round effects*						
Wald test	10.44	.316	4.08	.906	7.46	.589
σ_{RE} (sd.)	2.451 (.454)		1.301 (.241)		3.170 (.629)	
ρ (sd.)	.857 (.045)		.628 (.086)		.909 (.032)	

Note. Probit models on the probability of playing R estimated separately for each group of cognitive ability measured through the Raven test. For each group, models are estimated on pooled Human and Robot treatments data across all payoff configurations. (Exogeneous) unobserved individual heterogeneity is accounted for through random individual effects. The covariates are: round fixed effects, Robot treatment dummy and interactions between the two.

References

Abadie, A. (2002): “Bootstrap Tests for Distributional Treatment Effects in Instrumental Variable Model,” *Journal of the American Statistical Association*, 97(457), 284–292.

Akiyama, E., N. Hanaki, and R. Ishikawa (2015): “It is not just confusion! Strategic uncertainty in an experimental asset market,” *Economic Journal*, forthcoming.

Baron-Cohen, S., S. Wheelwright, J. Hill, Y. Raste, and I. Plumb (2001): “The “Reading the Mind in the Eyes" test revised version: A study with normal adults, and adults with Asperger syndrome or high-functioning autism,” *Journal of Child Psychology and Psychiatry*, 42(2), 241–251.

Battalio, R., L. Samuelson, and J. Van Huyck (2001): “Optimization Incentives and Coordination Failure in Laboratory Stag Hunt Games,” *Econometrica*, 69(3), 749–64.

Beard, T. R., and J. Beil, Richard O. (1994): “Do People Rely on the Self-Interested Maximization of Others? An Experimental Test,” *Management Science*, 40(2), 252–262.

Beard, T. R., R. O. J. Beil, and Y. Mataga (2001): “Reliant behavior in the United States and Japan,” *Economic Inquiry*, 39(2), 270–279.

Brañas-Garza, P., T. García-Muñoz, and R. Hernán (2012): “Cognitive effort in the Beauty Contest Game,” *Journal of Economic Behavior and Organization*, 83(2), 254–260.

Brandts, J., D. J. Cooper, and R. A. Weber (2014): “Legitimacy, Communication and Leadership in the Turnaround Game,” *Management Science*, forthcoming.

Burks, S. V., J. P. Carpenter, L. Goette, and A. Rustichini (2009): “Cognitive skills affect economic preferences, strategic behavior, and job attachment,” *Proceedings of the National Academy of Science*, 106(19), 7745–7750.

Camerer, C. F. (2003): *Behavioral Game Theory: Experiments in Strategic Interaction*. Princeton University Press, Princeton, NJ.

Carpenter, J., M. Graham, and J. Wolf (2013): “Cognitive ability and strategic sophistication,” *Games and Economic Behavior*, 80, 115–130.

Cooper, D. J., and J. B. Van Huyck (2003): “Evidence on the equivalence of the strategic and extensive form representation of games,” *Journal of Economic Theory*, 110(2), 290–308.

Cooper, R., and A. John (1988): "Coordinating Coordination Failures in Keynesian Models," *Quarterly Journal of Economics*, 103(3), 441–463.

Costa-Gomes, M. A., and V. P. Crawford (2006): "Cognition and Behavior in Two-Person Guessing Games: An Experimental Study," *American Economic Review*, 96(5), 1737–1768.

Davidson, R., and J. G. MacKinnon (2006): "Bootstrap methods in econometrics," in *Palgrave Handbooks of Econometrics*, ed. by T. C. Mills, and K. D. Patterson, vol. 1, chap. 23, pp. 812–838. Palgrave Macmillan, Basingstone.

Devetag, G., and A. Ortmann (2007): "When and why? A critical survey on coordination failure in the laboratory," *Experimental Economics*, 10(3), 331–344.

——— (2010): "Classic coordination failures revisited: the effects of deviation costs and loss avoidance," *Economics Bulletin*, 30(2), 1633–1641.

Dohmen, T., A. Falk, D. Huffman, and U. Sunde (2010): "Are Risk Aversion and Impatience Related to Cognitive Ability?," *American Economic Review*, 100(3), 1238–60.

Ellingsen, T., and R. Östling (2010): "When does communication improve coordination?," *American Economic Review*, 100(4), 1695–1724.

Embrey, M. S., G. R. Fréchette, and S. F. Lehrer (2014): "Bargaining and Reputation: An Experiments on Bargaining in the Presence of Behavioral Types," *Review of Economic Studies*, Forthcoming.

Fehr, D., and S. Huck (2015): "Who knows it is a game? On strategic awareness and cognitive ability," *Experimental Economics*, forthcoming.

Fehr, E., and J.-R. Tyran (2001): "Does Money Illusion Matter?," *American Economic Review*, 91(5), 1239–1262.

Ferraro, P. J., and C. A. Vossler (2010): "The Source and Significance of Confusion in Public Goods Experiments," *B.E. Journal of Economic Analysis & Policy*, 10(1), 1–42.

Fischbacher, U. (2007): "z-Tree: Zurich toolbox for ready-made economic experiments," *Experimental Economics*, 10(2), 171–178.

Frederick, S. (2005): "Cognitive Reflection and Decision Making," *Journal of Economic Perspectives*, 19(4), 25–42.

Georganas, S., P. J. Healy, and R. A. Weber (2015): "On the persistence of strategic sophistication," *Journal of Economic Theory*, 159(Part A), 369 – 400.

Gill, D., and V. Prowse (2015): "Cognitive ability, character skills, and learning to play equilibrium: A level-k analysis," *Journal of Political Economy*, forthcoming.

Goeree, J. K., and C. A. Holt (2001): "Ten Little Treasures of Game Theory and Ten Intuitive Contradictions," *American Economic Review*, 91(5), 1402–1422.

——— (2005): "An experimental study of costly coordination," *Games and Economic Behavior*, 51(2), 349–364.

Greiner, B. (2004): "An Online Recruitment System for Economic Experiments.," *University of Cologne, Working Paper Series in Economics*, 10, 79–93.

Heinemann, F. (2012): "Understanding Financial Crises : The Contribution of Experimental Economics," *Annales d'Economie et de Statistique*, 107-108, 7–29.

Houser, D., and R. Kurzban (2002): "Revisiting Kindness and Confusion in Public Goods Experiments," *American Economic Review*, 92(4), 1062–1069.

Ibanez, A., D. Huepe, R. Gempp, V. Gutiérrez, A. Rivera-Rei, and M. I. Toledo (2013): "Empathy, sex and fluid intelligence as predictors of theory of mind," *Personality and Individual Differences*, 54(5), 616–621.

Ivanov, A., D. Levin, and M. Niederle (2010): "Can relaxation of beliefs raiontalize the winner's curse?: An experimental study," *Econometrica*, 78(4), 1435–1452.

Jacquemet, N., R.-V. Joule, S. Luchini, and J. Shogren (2013): "Preference Elicitation under Oath," *Journal of Environmental Economics and Management*, 65(1), 110–132.

Jacquemet, N., and A. Zylbersztejn (2013): "Learning, words and actions: experimental evidence on coordination-improving information," *B.E. Journal of Theoretical Economics*, 13(1), 215–247.

Jacquemet, N., and A. Zylbersztejn (2014): "What drives failure to maximize payoffs in the lab? A test of the inequality aversion hypothesis," *Review of Economic Design*, 18(4), 243–264.

Johnson, E. J., C. Camerer, S. Sen, and T. Rymon (2002): "Detecting Failures of Backward Induction: Monitoring Information Search in Sequential Bargaining," *Journal of Economic Theory*, 104(1), 16–47.

Kamecke, U. (1997): "Rotations: Matching Schemes that Efficiently Preserve the Best Reply Structure of a One Shot Game," *International Journal of Game Theory*, 26(3), 409–417.

Morris, S., and H. S. Shin (2003): "Global Games: Theory and Applications," in *Advances in Economics and Econometrics*, ed. by M. Dewatripont, L. Hansen, and S. Turnovsky, chap. 3. Cambridge University Press.

Mullainathan, S., and E. Shafir (2013): *Scarcity: Why Having Too Little Means So Much.* Times Books, Henry Holt and Company, LLC, New York, NY.

Polonio, L., S. Di Guida, and G. Coricelli (2014): "Strategic sophistication and attention in games: an eye-tracking study," ECARES working paper 2014-22.

Raven, J. (2000): "The Raven's Progressive Matrices: Change and Stability over Culture and Time," *Cognitive Psychology*, 41(1), 1–48.

——— (2008): "The Raven progressive matrices tests: their theoretical basis and measurement model," in *Uses and abuses of Intelligence. Studies advancing Spearman and Raven's quest for non-arbitrary metrics*, ed. by J. Raven, and J. Raven, chap. 1, pp. 17–68. Cambridge University Press.

Rosenthal, R. W. (1981): "Games of perfect information, predatory pricing and the chain-store paradox," *Journal of Economic Theory*, 25(1), 92–100.

Sekhon, J. (2011): "Multivariate and Propensity Score Matching Software with Automated Balance Optimization," *Journal of Statistical Software*, 42(7), 1–52.

Selten, R. (1975): "Reexamination of the perfectness concept for equilibrium points in extensive games," *International Journal of Game Theory*, 4(1), 25–55.

Smith, V. L., G. L. Suchanek, and A. W. Williams (1988): "Bubbles, Crashes. and Endogenous Expectations in Experimental Spot Asset Markets," *Econometrica*, 56(5), 1119–1151.

Zeiliger, R. (2000): "A presentation of Regate, Internet based Software for Experimental Economics," *http://regate-ng.gate.cnrs.fr/ sferriol/.*